Continuous Ultrasound Speckle Tracking with Gaussian Mixtures

Colas Schretter[1,2], Jianyong Sun[3], Shaun Bundervoet[1,2], Ann Dooms[1,2], Peter Schelkens[1,2]
Catarina de Brito Carvalho[2,4], Pieter Slagmolen[2,4] and Jan D'hooge[2,5]

***Abstract*— Speckle tracking echocardiography (STE) is now widely used for measuring strain, deformations, and motion in cardiology. STE involves three successive steps: acquisition of individual frames, speckle detection, and image registration using speckles as landmarks. This work proposes to avoid explicit detection and registration by representing dynamic ultrasound images as sparse collections of moving Gaussian elements in the continuous joint space-time space. Individual speckles or local clusters of speckles are approximated by a single multivariate Gaussian kernel with associated linear trajectory over a short time span. A hierarchical tree-structured model is fitted to sampled input data such that predicted image estimates can be retrieved by regression after reconstruction, allowing a (bias-variance) trade-off between model complexity and image resolution. The inverse image reconstruction problem is solved with an online Bayesian statistical estimation algorithm. Experiments on clinical data could estimate subtle sub-pixel accurate motion that is difficult to capture with frame-to-frame elastic image registration techniques.**

***Index Terms*— Speckle tracking echocardiography (STE), Gaussian mixture model (GMM), local motion estimation**

I. Introduction

Speckle tracking echocardiography (STE) [1] is a motion capture method that emerged from cardiology and is now in the midst of being translated to new medical applications such as measuring strain and stress of small structural tendons and ligaments in orthopedics [2]. Image-guided surgical applications are also foreseen by correlating dynamic image observations with advanced model-based biomechanics simulations of deformation [3].

In STE, motion estimation can be obtained with elastic image registration methods that use strong regularization priors to ensure spatial smoothness and local invertibility of the deformation vector field [4]. As a consequence, these techniques are not yet capable of capturing local motion occurring in small scale soft tissues, despite their robustness [5]. This work proposes to bridge this gap by avoiding computing explicit registration. Instead, the dynamic ultrasound image and motion of speckles is represented in a single unified spatio-temporal model. Estimated local linear motion trajectories are associated to each moving image element and can be retrieved from the model retrospectively.

This work is part of the iMinds ICON 3DUS project (Strain quantification using 3D UltraSound imaging for musculoskeletal applications). J. Sun is supported by The Key Science and Technology Project of WuHan under Grant No. 2014010202010108.

[1]Dept. of Electronics and Informatics (ETRO), Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. [2]iMinds, Gaston Crommenlaan 8, box 102, 9050 Ghent, Belgium. [3]The University of Greenwich, Faculty of Engineering and Science, UK. [4]Medical Imaging Research Center, UZ Herestraat 49 - box 7003, 3000 Leuven, Belgium. [5]Cardiovascular Imaging and Dynamics, UZ Herestraat 49 - box 7003, 3000 Leuven, Belgium.

We represent dynamic ultrasound images as a multi-resolution mixture of weighted multivariate Gaussians in 2D+T dimensions, as illustrated in Figure 1. The image and motion is therefore stored explicitly at varying levels of details in the scale-space [6]. Starting from uniform initialization, maximum likelihood values of parameters are estimated by an online conditional expectation-maximization method [7]. After parametric estimation, the subtle motion of speckles is captured in the sparse model, using few mixture components. The rationale of our approach is that solving together dynamic image formation and motion estimation as a single problem may be better conditioned than performing successively frame-by-frame speckle identifications and tracking from the ultrasound image sequence.

The remainder of this paper is structured as follows. In section II, we define the input image data and a continuous statistical model to represent joint image and motion. Section III describes a Bayesian method for estimating maximum-likelihood values of model's parameters. In section IV, we discuss results from experiments on an acquisition of a clamped and stretched tendon. Finally, we conclude and pinpoint next research steps in section V.

II. Data and Image Models

We collect temporally apodized raw B-mode images in a discrete spatio-temporal amplitude field $A(x,y,t)$, with $x \in \{1,\ldots,N_x\}$, $y \in \{1,\ldots,N_y\}$ and $t \in \{1,\ldots,N_t\}$. This 2D+T matrix of real positive amplitude values is first normalized for every time frame t such that

$$\sum_{x=1}^{N_x}\sum_{y=1}^{N_y} A(x,y,t) = 1,\ \forall t \in \{1,\ldots,N_t\}.$$

We approximate $A(x,y,t)$ with a continuous and smooth Gaussian mixture model (GMM) $G(x,y,t)$ having the form

$$A(x,y,t) \approx G(x,y,t) = \sum_{k=1}^{K} w_k \, \frac{g\left(x,y,t;\mu_k,\Sigma_k\right)}{g\left(t;\mu_k,\Sigma_k\right)},$$

with the set of component's weights, means, and symmetric variances-covariances matrices

$$\Theta = \left\{ w_k, \mu_k = \begin{bmatrix}\mu_x\\ \mu_y\\ \mu_t\end{bmatrix}, \Sigma_k = \begin{bmatrix}\sigma_x^2 & \sigma_{xy} & \sigma_{xt}\\ \sigma_{xy} & \sigma_y^2 & \sigma_{yt}\\ \sigma_{xt} & \sigma_{yt} & \sigma_t^2\end{bmatrix} \right\}_{k=1}^{K}$$

where $g\left(x,y,t;\mu_k,\Sigma_k\right)$ is the product of a 1D marginal (temporal) and the 2D conditional (spatial) distributions:

$$g\left(x,y,t;\mu_k,\Sigma_k\right) = g\left(t;\mu_t,\sigma_t^2\right) \times g\left(x,y;\bar{\mu},\bar{\Sigma}\right),$$

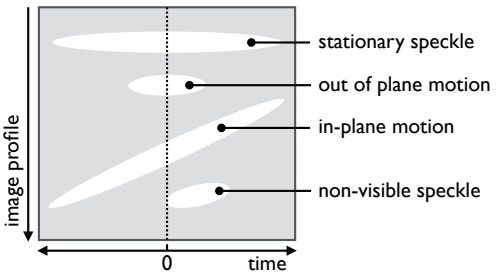


Fig. 1. Diagram of various types of temporal correlations of speckle positions that can be captured by the joint space-time Gaussian mixture model. Predictions of 2D frames represented by the dotted vertical line profile are retrieved from conditional regression given any positive or negative offset from the reference time frame.

with the mean of the bivariate conditional distribution

$$\bar{\mu} = \begin{bmatrix}\mu_x \\ \mu_y\end{bmatrix} + \begin{bmatrix}\sigma_{xt} \\ \sigma_{yt}\end{bmatrix} (t - \mu_t)/\sigma_t^2$$

and the conditional variances-covariance matrix

$$\bar{\Sigma} = \begin{bmatrix}\sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2\end{bmatrix} - \begin{bmatrix}\sigma_{xt} \\ \sigma_{yt}\end{bmatrix} \begin{bmatrix}\sigma_{xt} & \sigma_{yt}\end{bmatrix} / \sigma_t^2 \,.$$

The univariate Gaussian kernel $g\left(x; \mu, \sigma^2\right) \equiv g_1$ is

$$g_1 = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right],$$

and the bivariate Gaussian kernel $g\left(x, y; \mu, \Sigma\right) \equiv g_2$ is

$$g_2 = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left[-\frac{1}{2}\left(\begin{bmatrix}x \\ y\end{bmatrix} - \mu\right)^\top \Sigma^{-1} \left(\begin{bmatrix}x \\ y\end{bmatrix} - \mu\right)\right].$$

III. Reconstruction

For estimating the mixture parameters Θ for the given reference time t_r, we first multiply 2D frames from the input sequence A with $g\left(t; t_r, \sigma^2\right)$ with σ set to half the number of consecutive frames where the motion is assumed to be linear. After apodization, we sample non-uniformly points in the domain of $A(x, y, t)$ by generating a source sequence

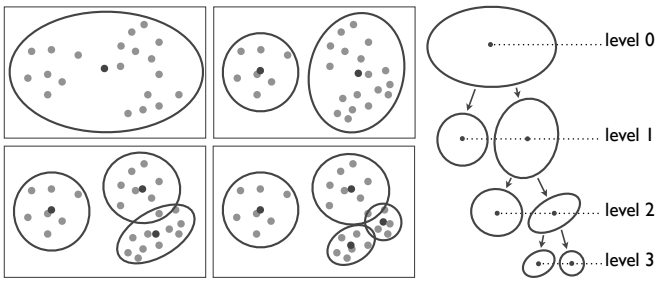
$$D = \{(x_m, y_m, t_m)\}_{m=1}^M$$

such that samples are drawn from the apodized 2D+T discrete probability density function A. We use a single inversion method for non-uniform sampling [8].

The set Θ of all weights and parameters of the mixture $G(x, y, t)$ is now estimated by maximizing the log-likelihood

$$P(D; \Theta) = \sum_{m=1}^{M} \log \sum_{k=1}^{K} w_k \, \frac{g\left(x_m, y_m, t_m; \mu_k, \Sigma_k\right)}{g\left(t_m; \mu_k, \Sigma_k\right)} \,.$$

We use an online expectation-maximization (EM) method [7] for estimating the model parameters. For each data sample, EM alternates between performing an expectation step (E) which evaluates membership probabilities using current model parameters, and a maximization step (M), which updates parameters for increasing the data likelihood.



Multi-resolution subdivision Tree structure

Fig. 2. Example of image formation from few data samples marked by grey discs. Four successive levels of a binary hierarchy are shown in frames and the multi-resolution tree structure is drawn on the right. The iso-contours of multivariate Gaussians show that each component approximates a local subset of the input data.

A. Online expectation-maximization

In the E-step, we evaluate the model from the information carried by the m-th sample (x_m, y_m, t_m) by computing the membership probability $p_k(x_m, y_m | t_m)$, expressing the probability that the sample is drawn from the conditional distribution of the component k at given time t_m using

$$p_k(x_m, y_m | t_m) = \frac{w_k \; g\left(x_m, y_m; \bar{\mu}, \bar{\Sigma}\right)}{\sum_{k=1}^{K} w_k \; g\left(x_m, y_m; \bar{\mu}, \bar{\Sigma}\right)} \,,$$

with the parameters $\bar{\mu}$ and $\bar{\Sigma}$ of the conditional distribution.

In the second M-step, the weight w_k is incremented for each $k \in \{1, \ldots, K\}$ with $w_k \leftarrow w_k + w_k \, \alpha$ and kernel parameters μ_k and Σ_k are updated such that the likelihood of observing the new sample is improved as follows:

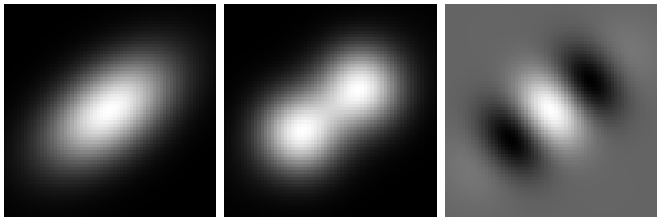
$$\mu_k \leftarrow \mu_k + \delta_k \, \alpha \quad \text{and} \quad \Sigma_k \leftarrow \left[\Sigma_k + \delta_k \, \delta_k^\top \, \alpha\right] (1 - \alpha)$$

with $\alpha = p_k(x_m, y_m | t_m)$ and $\delta_k = \begin{bmatrix}x_m & y_m & t_m\end{bmatrix}^\top - \mu_k$.

In order to adapt smoothly to new data, a mechanism should be used to adjust parameters to the most recent data samples. In order to guarantee that every data sample in a finite sliding window over the recent stream of data samples has equal importance, we stored the history of all incremental weight updates in a sliding window of given size $T = 256$. The value of T should be large enough to guarantee a stable statistical estimates of component's mean and covariance matrix. Therefore T is controling the bias-variance tradeoff. The circular history buffer is segmented in 64 memory pages that accumulate batches of successive updates. No significant convergence speed increase was observed with more pages.

B. Tree-structured initialization

We use a multi-resolution binary tree structure as shown in Figure 2 to represent the mixture model [9] instead of choosing *a priori* the number of mixture component K and initializing parameters with *ad hoc* values that are not specifically dependent on the input data. For creating the tree hierarchy, the model is initialized with $K = 1$, and maximum-likelihood parameters are computed for this single component using the first data samples from D.



Original kernel Binary splitting Difference

Fig. 3. Binary splitting operation on a 2D Gaussian kernel (left), giving two displaced and squashed children (middle). The positive (bright) and negative (dark) differences integrate to zero (right).

Once a mixture component k is supported by more that $T = 256$ data samples, the conditional Gaussian kernel is split in two and siblings are displaced in opposite direction along its normalized dominant eigenvector $e_k = \begin{bmatrix} e_x & e_y \end{bmatrix}^\top$ corresponding to the principal eigenvalue λ_k w.r.t. Σ_k. Figure 3 shows an example of this binary splitting procedure.

The eigenvector e_k is aligned with the longest axis of the first standard deviation contour ellipse and λ_k is the squared radius. Using the orthogonal squashing operator

$$Q = \begin{bmatrix} 1 - \frac{3}{4}e_x^2 & -\frac{3}{4}e_x\, e_y \\ -\frac{3}{4}e_x\, e_y & 1 - \frac{3}{4}e_y^2 \end{bmatrix} \quad \text{and} \quad \Sigma_{xy} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix},$$

we modify the variances-covariances matrices of children:

$$\Sigma_{2k} = \Sigma_{2k+1} = \begin{bmatrix} S_{1,1} & S_{1,2} & s_{xt} \\ S_{2,1} & S_{2,2} & s_{yt} \\ s_{xt} & s_{yt} & \sigma_t^2 \end{bmatrix},$$

with $S = Q\Sigma_{xy}$ and $\begin{bmatrix} s_{xt} & s_{yt} \end{bmatrix}^\top = Q \begin{bmatrix} \sigma_{xt} & \sigma_{yt} \end{bmatrix}^\top$.

Then, we displace centers of the two new components with

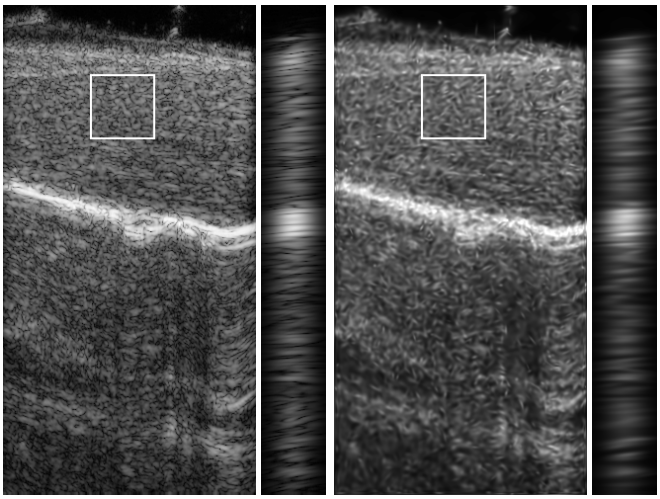
$$\delta = \left[e_x \quad e_y \quad \begin{bmatrix} \sigma_{xt} & \sigma_{yt} \end{bmatrix} \Sigma_{xy}^{-1} \begin{bmatrix} e_x & e_y \end{bmatrix}^\top \right]^\top \sqrt{\tfrac{3}{4}\lambda_k}$$

such that $\mu_{2k} = \mu_k + \delta$ and $\mu_{2k+1} = \mu_k - \delta$. Finally, we share the weight in two: $w_{2k} = w_{2k+1} = w_k/2$.

As we can see, the indices correspond to the storage of a binary tree in a vector, with breadth-first traversal. After convergence, the leaf components provides a finely-resolved image model. We also keep the intermediate nodes in a binary tree for recovering models at lower resolution levels. Since the image is made of Gaussian components, the tree is akin to a multi-resolution scale-space representation [6].

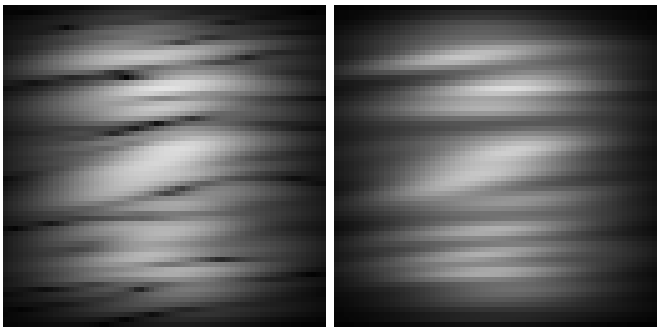
IV. Experiments

We conducted experiments on an 2D+T acquisition of an isolated digital flexor tendon of a sheep using a Vevo 2100 ultrasound machine (FUJIFILM VisualSonics, Inc, Toronto, CA) with a 256 elements linear array transducer. The central frequency was 20MHz and images were acquired at 50 frames per second. The tendon was mechanically stretched on a loading platform in a continuous strain between 2% to 6%. The apparent motion of speckles is relatively smooth in the sequence of 128 frames. Since speckles are intrinsically three-dimensional shapes and only a planar cut over time is captured, the out of plane motion translates into smooth fading in and out of 2D cross sections through speckles.



Original reference frame Model regression ($K = 7936$)

Fig. 4. Reference frame and optical flow images of the central line profile extracted from the raw input sequence of ultrasound B-mode frames (left) as well as a high-precision Gaussian mixture model reconstruction using 7936 components. White squares delineate the region of interest of 64 pixels that is used for close-up illustrations.

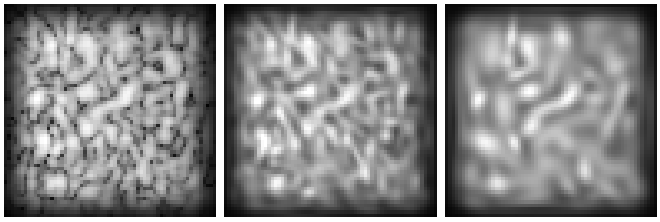


Reference optical flow Model regression ($K = 256$)

Fig. 5. Optical flow images of the central image line profile extracted from the ground truth input data and the regression from the estimated Gaussian mixture model. The multivariate Gaussians fit moving speckles that form the input sequence. The linear motion prediction model is in good agreement with reference observations.

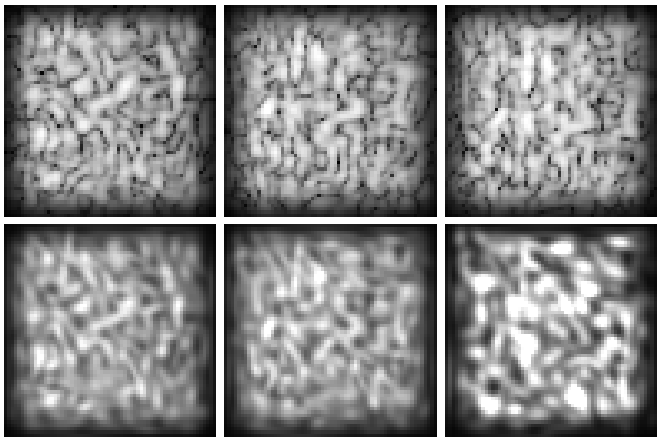
The input sequence of B-mode images contains 128 frames of 496×256 pixels and the motion of speckles is visible from subtle image intensity changes. In order to meet the local linear motion hypothesis, the input dataset is smoothly apodized over time using a Gaussian kernel at the central reference frame 64 with standard deviation $\sigma = 32$. This parameter is trading-off the temporal resolution of the motion model and the robustness (or stability) of the statistical motion estimate. The effect of apodization weighting can be seen in the optical flows in Figure 4. We reconstructed the whole dataset with $K = 7936$ components that correspond to a density of 256 components for the close-up 64×64 pixels window used for optical flows in Figure 5.

Figure 6 shows the effect of the model complexity on the smoothness of the image approximation. Selecting a small number of components regularizes the estimation problem



Reference Model ($K = 256$) Model ($K = 64$)

Fig. 6. Progressive image resolution improvement with the number of mixture components K. An optimal agreement with data is obtained when K is close to the number of visible speckles in the reference B-mode image frame. However, a much smaller number of components is often sufficient for smooth motion estimation.



Offset: -16 Offset: $+32$ Offset: $+64$

Fig. 7. Prediction of speckle's position with time varying time offsets from the reference frame. The ground truth reference frames are shown on the top row and the predictions from the Gaussian mixture model with $K = 256$ components are on the bottom row.

in a natural way. We observed that around 256 mixture components per region of 64×64 pixels was satisfactory for capturing motion of speckles. In contrast to pixelized images, linear trajectories of speckle image elements are explicitly stored in the covariance parameters of variances-covariances matrices that are estimated individually for each component.

Figure 7 illustrates the precision of predicted frames using linear displacements of speckles that are identified in the reference frame. For an offset until 32 frames, the smooth drift of speckles is still successfully predicted. With larger offsets, linear trajectories are crossing and spurious clusters appear. Prediction accuracy drops when the hypothesis of local linear trajectories does not hold anymore. In Figure 8, we compared line profiles in the ground truth frames with the prediction of past and future frames using the sparse Gaussian mixture. Sub-pixel accurate relative displacements are measured from the horizontal shift of reference landmarks in the center of speckle bumps. These results on clinical data confirm our first study on 1D profiles from a numerical phantom [10].

V. Conclusion

We represent dynamic ultrasound images as a mixture of moving smooth ellipsoids that are approximated by multivariate Gaussian kernels. Bayesian parametric estimation is carried out a fast online expectation-maximization method and locally linear approximations of speckle's motion are retrieved with conditional regression from the model. The potential of this new image representation for motion estimation in ultrasound speckle tracking has been experimentally verified on a clinical acquisition of an elongated tendon mounted in a motorized clamper.

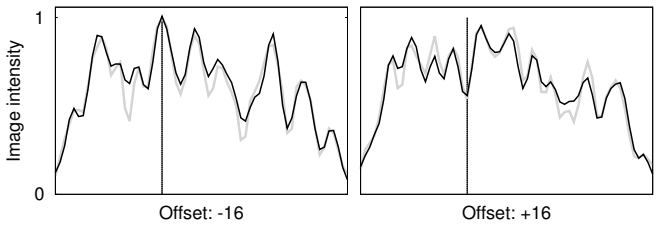


Fig. 8. B-mode line profiles extracted from the reference input frames (gray) and the 16 frames in past and future predictions from conditional Gaussian mixture model regression (black).

Sparse mixture models could reconstruct fair approximations of the smooth motion vector field. The running time complexity depends on the amount of data that must be sampled before obtaining a satisfactory convergence of the statistical estimation method. Only mixture components that are close to the data sample must be updated ass well. Therefore, there is an advantage of using the tree structure for accelerating substantially computations by pruning branches that do not overlap significantly with the data sample or the query location during regression. Future work will include further evaluations of running time requirements and quantitative comparisons of motion flows on digital phantoms.

References

[1] J. D'hooge, A. Heimdal, F. Jamal *et al.*, "Regional strain and strain rate measurements by cardiac ultrasound: principles, implementation and limitations," *Europ. J. of Echocardiography*, pp. 154–170, 2000.

[2] L. A. Chernak and D. G. Thelen, "Tendon motion and strain patterns evaluated with two-dimensional ultrasound elastography," *J. of Biomechanics*, vol. 45, no. 15, pp. 2618–2623, 2012.

[3] T. S. Pheiffer, R. C. Thompson, and D. Rucker, "Model-based correction of tissue compression for tracked ultrasound in soft tissue image-guided surgery," *Ultrasound in Medicine & Biology*, no. 4, pp. 788–803, 2014.

[4] S. Y. Chun, C. Schretter, and J. A. Fessler, "Sufficient condition for local invertibility of spatio-temporal 4D B-spline deformations," in *Proc. of the IEEE ISBI*, 2010, pp. 1221–1224.

[5] B. Heyde, R. Jasaityte, D. Barbosa *et al.*, "Elastic image registration versus speckle tracking for 2-D myocardial motion estimation: A direct comparison in vivo," *IEEE Transactions on Medical Imaging*, vol. 32, no. 2, pp. 449–459, 2013.

[6] T. Lindeberg, "Scale-space theory: A basic tool for analysing structures at different scales," *J. of Applied Statistics*, vol. 21, pp. 224–270, 1994.

[7] R. M. Neal and G. E. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in Graphical Models*. MIT Press, 1998, pp. 355–368.

[8] C. Schretter and H. Niederreiter, "A direct inversion method for non-uniform quasi-random point sequences," *Monte Carlo Methods and Applications*, vol. 19, no. 1, pp. 1–9, 2013.

[9] V. Garcia, F. Nielsen, and R. Nock, "Hierarchical gaussian mixture model." in *ICASSP*, 2010, pp. 4070–4073.

[10] S. Bundervoet, C. Schretter, A. Dooms, and P. Schelkens, "Bayesian estimation of sparse smooth speckle shape models for motion tracking in medical ultrasound," in *iTWIST'14, Namur, Belgium*, 2014, pp. 7–9.